

Informing about Web Paradata Collection and Use (Version 1.0)

Kunz, Tanja; Beuthner, Christoph; Hadler, Patricia; Roßmann, Joss;
Schaurer, Ines

Veröffentlichungsversion / Published Version

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Kunz, T., Beuthner, C., Hadler, P., Roßmann, J., & Schaurer, I. (2020). *Informing about Web Paradata Collection and Use (Version 1.0)*. (GESIS Survey Guidelines). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/gesis-sg_036

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-nc/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more Information see:
<https://creativecommons.org/licenses/by-nc/4.0>

Informing about Web Paradata Collection and Use

*Kunz, Tanja, Christoph Beuthner, Patricia Hadler, Joss Roßmann, and
Ines Schaurer*

Abstract

This survey guideline addresses the practical question of how best to inform survey participants about the collection and use of paradata in web surveys. We provide an overview of different personal and non-personal web paradata and the associated information and consent requirements. Best practices regarding the procedure, wording, and placement of non-personal web paradata information are discussed. In addition, we propose a sample wording for web paradata information in German and English.

Citation

Kunz, Tanja, Christoph Beuthner, Patricia Hadler, Joss Roßmann, and Ines Schaurer (2020). *Informing about Web Paradata Collection and Use*. Mannheim, GESIS – Leibniz-Institute for the Social Sciences (GESIS – Survey Guidelines).

DOI: 10.15465/gesis-sg_en_036

This work is licensed under a Creative Commons Attribution – NonCommercial 4.0 International License (CC BY-NC).



1. Introduction

In survey research, paradata generally refer to information describing the process of survey data collection. In web surveys, paradata are primarily captured automatically as a by-product to help understand and improve the survey data collection process (Couper, 1998; Kreuter, 2013).

Given changing data protection regulations (i.e., the General Data Protection Regulation [GDPR] applicable as of May 25th, 2018 in the EU) and the continuous development of paradata scripts to gather client-side paradata in web surveys (Heerwegh, 2003; Kaczmirek & Neubarth, 2007; Schlosser & Höhne, 2020; see also the GESIS Survey Guideline “Collection and Use of Web Paradata” by Kunz & Hadler, 2020), an ongoing discussion has started on the need to inform the respondents about the collection and use of their paradata by professional associations in the field of market, opinion and social research (ADM, ASI, BVM, & DGOF, 2007; ESOMAR/GRBN, 2015, 2017) as well as in the scientific community (e.g., Couper, 2017; Felderer & Blom, 2019). So far, however, existing ethical codes and guidelines in survey research generally remain vague on this issue.

This guideline discusses current information and consent requirements and guides how best to inform survey participants about the collection and use of web paradata. We focus our recommendations on paradata which are collected in scientific web surveys with voluntary participation of individuals and fall into the category of non-personal paradata. Although we refer in many parts to the European data protection regulation, national interpretations may differ. The recommendations we make therefore refer to web (para-)data collected in Germany. In this context, we point out that legal requirements for informed consent can change, as can the interpretation of the general guidelines in survey research. All statements and recommendations made here are subject to any changes in legal regulations or the ethical principles applicable to survey research. Although the authors have presented their knowledge in the best possible way, they do not assume any guarantees regarding the legal validity of the statements made. This guideline is not to be understood as binding legal advice. We point out that it is always the researchers’ responsibility to check the framework conditions applicable to their project and to inform themselves about changes in the legal and ethical requirements before conducting a survey. A consultation with the data protection officer of the institution regarding the consent requirements for the collection and use of web paradata is strongly recommended.

2. Background

2.1 Legal requirements and research ethics

The General Data Protection Regulation [GDPR] is the European Union regulation that harmonizes the rules for processing personal data throughout the EU (see <https://gdpr-info.eu/>).

Personal data (or *personally identifiable information*, PII) means “any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” (see Article 4(1), GDPR).

Before collecting and processing *personal data*, the respondents' active and informed consent needs to be sought. 'Informed consent' is the most commonly used lawful exception for the collection and processing of this type of data (see Article 6, GDPR). If used, it must be obtained through affirmative action (i.e., opt-in procedure; see also section 2.2). According to Article 4(11), GDPR, "consent of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by clear affirmative action, signifies agreement to the processing of personal data relating to him or her." Researchers must provide a detailed description of the personal data they intend to collect, the purposes for which the data are processed, and further data processing steps. Respondents, in turn, must actively agree to the collection and use of their personal data.¹

If the purpose of processing is scientific research, however, there are some legal exceptions concerning collecting and using personal data (see Article 89, GDPR). Similar legislation exists in Germany, which allows the collection and use of personal data for scientific research without prior consent under limited circumstances. These exceptions are specified in Article 27 of the Federal Data Protection Act (BDSG-neu)², addressing "data processing for purposes of scientific or historical research and for statistical purposes." However, Article 27, BDSG-neu in no way releases the researcher from his or her obligation to provide information on collecting and using personal data in advance of the survey data collection. Moreover, data subjects retain the right to erasure their personal data (Article 17, GDPR).

The existing ethical codes and guidelines of professional associations addressing legal and ethical considerations in survey research are generally not clear on information and consent requirements related to web paradata. For example, the joint guideline for online surveys published by the professional associations ADM, ASI, BVM, and DGOF (2007) in Germany recommends that in cases where *additional* data are to be *imperceptibly collected and stored* during a web survey, respondents should be asked for their consent in advance (please note that this guideline is currently being revised). There are few exceptions where the guidelines advocate active consent for web paradata on ethical grounds. One example is keyloggers³ to monitor and record each keystroke typed on a keyboard (ESOMAR/GRBN, 2015, p. 24).

2.2 Types of consent

Consent means the "freely given and informed indication of agreement by a person to the collection and processing of his/her personal data" (ESOMAR/GRBN, 2015, p. 6). We can distinguish between two types of *explicit consent* (see Table 1). There are *active* (opt-in) consent procedures, where respondents must actively consent to data collection and use (e.g., ticking a box indicating consent). And *passive* (opt-out) consent procedures, where respondents must disagree with the request (e.g., by clicking a button to withdraw consent) (Sakshaug, Schmucker,

¹ Further details on the conditions for consent are detailed in Article 7 and Recitals 32 and 42 of the GDPR. Further information can also be found in the data protection guideline of the German Data Forum (RatSWD, 2020). For further information, see Schaar (2017) and RatSWD (2020). Moreover, researchers are advised to consult the GDPR (<https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04>) with scrutiny concerning consent requirements (<https://gdpr.eu/gdpr-consent-requirements/>). The GDPR website offers a checklist for consent to the collection of personal data (<https://gdpr.eu/checklist/>), as well as templates for data processing agreements (<https://gdpr.eu/data-processing-agreement/>) and right to erasure (<https://gdpr.eu/right-to-erasure-request-form/>).

² The BDSG-neu supplements and concretizes the specifications of the GDPR without contradicting them.

³ Keyloggers can be collected with most open-source, client-side paradata scripts (e.g., UCSP by Kaczmirek & Neubarth, 2007; ECSP by Schlosser & Höhne, 2020). However, researchers are free to choose not to collect all paradata types possible with these scripts. Therefore, researchers who do not need keyloggers for analysis should not implement this part of the script.

Kreuter, Couper, & Singer, 2016). A discussion of the advantages and disadvantages of each procedure is given in Sakshaug et al. (2016).

Merely informing respondents of the intention to collect and use their (para-)data without explicitly asking for their consent—sometimes called implicit or implied consent—is strictly speaking, not a form of informed consent. It means that respondents are informed about the collection and use of their data (e.g., on the welcome page of a web survey); their subsequent participation in the (web) survey is *interpreted* as consent.

Table 1. Different types of information and consent procedures

	Definition
Explicit consent procedure	<p>The respondent must explicitly consent to the collection and use of data.</p> <ul style="list-style-type: none"> – <i>Active, opt-in consent procedure:</i> The respondent must actively agree to data collection and use (e.g., by checking a box indicating consent). – <i>Passive, opt-out consent procedure:</i> The respondent must actively disagree with data collection and use (e.g., by clicking a button to withdraw consent).
Information procedure (also referred to as implicit consent)	<p>The respondent implicitly agrees to the collection and use of the data by continuing with the survey after being informed about the collection and use of data.</p>

2.3 Types of web paradata and implications for information and consent requirements

We can distinguish between *personal* and *non-personal web paradata*, which involves different information and consent requirements. In the following, we use the term personal web paradata when referring to personally identifiable information, i.e., paradata that allow conclusions to be drawn about a person. Non-personal web paradata thus refer to non-identifying information, i.e., paradata that do not allow any conclusions about the person.

Personal web paradata

Web paradata as data about the process of collecting survey data (e.g., device type information, timestamps) are usually non-personal data. However, there are a few exceptions that are considered personal web paradata (e.g., geolocation). It should also be noted that a combination of different types of non-personal web paradata or a combination of non-personal web paradata and survey data may result in personally identifiable information (e.g., timestamps, device type, and socio-demographic information in an in-house survey).

In any case, survey researchers should always decide on a case-by-case basis whether personal paradata are involved, and if so, whether active opt-in consent is required or whether mere information is sufficient (e.g., if Article 27, BDSG-neu can be invoked; see above).

The following non-exhaustive list contains the most common types of personal web paradata, because they can also be considered personal data per se (see Article 4(1), GDPR):

- IP addresses as a unique identifier on the network.
Most survey software providers can provide IP addresses; although these are usually not collected by default, the researcher should check this in the settings before starting a survey.

- Geolocation information relating to “the identification of the real-world geographic location of an object, such as a computing device (computer, tablet, smartphone, etc.)” (ESOMAR/GRBN, 2015, p. 7).

Many survey software providers offer the option of having the geolocation of the participants via the IP address (“Standortinformationen der Teilnehmer über die IP-Adresse ermitteln”). By default, this data is not collected by most providers; however, researchers should verify this in the settings.

Non-personal web paradata

In survey research, most web paradata are non-personal information that does not allow any conclusions about a person. Table 2 gives an overview of different non-personal web paradata (in English and German language).

Under the current legal basis (i.e., GDPR and BDSG-neu), explicit consent for collecting and using non-personal web paradata is not required. However, from a research ethics’ perspective, it is often recommended that researchers inform their respondents about non-personal web paradata. Best practices for informing about the collection and use of non-personal web paradata are described in detail in the next section.

Table 2. Types of non-personal web paradata (in English [EN] and German [DE] language)

	Examples	Further subtypes
Non-personal (EN)	<ul style="list-style-type: none"> – Device information – Keyboard, mouse, and touch screen input – Date, time, and timestamp data 	<ul style="list-style-type: none"> Type of device, screen size, resolution and orientation, operating system, browser, Internet connection Keyboard inputs, mouse/finger clicks, mouse/finger movements, and mouse/finger positions, scrolling, leaving the survey Current time and date, survey completion time
Nicht-personenbezogen (DE)	<ul style="list-style-type: none"> – Angaben zum Endgerät – Tastatur-, Maus- und Touchscreeneingaben – Datums-, Zeit- und Zeitstempelangaben 	<ul style="list-style-type: none"> Art des Geräts, Bildschirmgröße, -auflösung und -ausrichtung, Betriebssystem, Browser, Internetverbindung Tastatureingaben, Maus-/Fingerklicks, Maus-/Fingerbewegungen und Maus-/Fingerpositionen, Scrollen, Verlassen der Umfrage Aktuelle Uhrzeit und Datum, Bearbeitungszeit

Whenever the collection and use of web paradata require respondent information or, in particular active opt-in consent, researchers should consider the possibility of reduced survey participation or other adverse effects on respondent behavior in their decision to collect web paradata at all. So far, there are only a few studies on informed consent for web paradata use. These studies show mixed results regarding survey participation and breakoff. Apart from this, almost no significant effects on response behavior (e.g., item nonresponse, straightlining, response time) were found (Couper & Singer, 2013; Kunz & Gummer, 2020; Kunz, Landesvatter, & Gummer, 2020; Sattelberger, 2015). Previous research on informed consent for information that is usually considered much more sensitive than web paradata (e.g., linking survey data to administrative data, collecting physical measurements, or tracking sensor data) shows that at least some respondents are then no longer willing to participate in the survey (Keusch, Struminskaya, Antoun, Couper, & Kreuter, 2019; Revilla, Couper, & Ochoa, 2019).

3. Web Paradata Information: Best Practice for Procedure, Wording, and Placement

In line with the ethical guidelines for survey research, we recommend that respondents are informed about the collection and use of their web paradata, even if active, opt-in consent is not mandatory (see section 2.2). According to the principle of transparency, the information should be made available “in a concise, transparent, intelligible and easily accessible form, using clear and plain language” (see Article 12, GDPR). Applied to web paradata information, this means that respondents should receive a clear description of the type of web paradata collected and the purposes for which they are used. Descriptions should be kept simple and written in a clear and understandable language. Giving examples can help to improve understanding. All information should be easily accessible to everyone. These general requirements concern the procedure, wording, and placement of web paradata information, which we describe in more detail below.

3.1 Procedure

If researchers inform their respondents about the collection and use of web paradata, the information should be provided before or at the beginning of the survey data collection. Respondents are thus informed before the survey that web paradata will be collected during survey participation and used during and/or after the end of the survey.

3.2 Wording

Web paradata information should generally address the following two issues:

1. What kind of paradata will be collected?

Following the principle of data minimization (see Article 5 (1) c, GDPR), it is recommended that the collection of data is “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.”

Regarding the collection and use of web paradata, the information should be provided about the types of web paradata collected by the survey software or, if applicable, by the additional paradata scripts implemented in the web survey.

Basic information about the different types of web paradata collected in a survey (e.g., device information) is usually sufficient. Optionally, detailed information in the form of a (complete) list of all subtypes of web paradata collected in the survey (e.g., type of device, screen size, operating system) can be provided. It is strongly recommended to always adapt the web paradata information to the current web survey.

2. What are the purposes of the processing?

In line with the purpose limitation principle (see, e.g., Article 5 (1) b, GDPR), researchers should ensure that data are “collected for specified, explicit and legitimate purposes.”

Applied to web paradata, researchers should also explain the purpose of their collection and use.

3.3 Placement

Web paradata information, including basic information about the type and purpose of the web paradata collected, is best provided on the welcome or start page of the web survey (usually the first webpage) to make the information easily accessible to all respondents. Optionally, detailed information that includes the full list of web paradata and other details about the collection and

use of web paradata may be provided via a link to the “Data Use and Privacy Policy” section of an external website or survey homepage. It is additionally recommended to provide a contact address (e.g., e-mail) on the first (and last) page for any questions regarding the survey in general or the collection and use of web paradata.

3.4 Sample wording

Table 3. Sample wording of web paradata information (in English [EN] and German [DE] language)

EN	<p>In addition to your answers, technical data and information on the course of this survey will also be collected and evaluated. This includes [time data], [keyboard], [mouse] [or] [touch screen input] [as well as] [information on the device you are using].</p> <p>This information is used exclusively for [scientific research purposes] [and] [the optimal presentation of the questionnaire on your device]. They are of great value for scientific research and help to improve surveys.</p> <p>The analysis of the information is, of course, carried out without drawing conclusions about your person.</p>
DE	<p>Neben Ihren Antworten werden auch technische Daten und Informationen zum Verlauf dieser Befragung erhoben und ausgewertet. Hierzu zählen [Zeitangaben], [Tastatur-], [Maus-] [oder] [Touchscreeneingaben] [sowie] [Angaben zu dem von Ihnen verwendeten Endgerät].</p> <p>Diese Informationen werden ausschließlich für [wissenschaftliche Forschungszwecke] [und] [die optimale Darstellung des Fragebogens auf Ihrem Gerät] verwendet. Sie sind für wissenschaftliche Untersuchungen von großem Wert und helfen, Befragungen zu verbessern.</p> <p>Die Auswertung der Informationen erfolgt selbstverständlich ohne Rückschlüsse auf Ihre Person.</p>

Note. Sample wording of web paradata information, including basic information about the type and purpose of the web paradata collected, that is best provided on the welcome or start page of the web survey. Information in parentheses is optional and should always be adapted to the type and purpose of the web paradata collected in the current web survey.

References

- ADM, ASI, BVM, & DGOF (2007). *Richtlinie für Online-Befragungen*. Retrieved from http://rat-marktforschung.de/fileadmin/user_upload/pdf/R08_RDMS.pdf
- Couper, M. P. (1998). Measuring Survey Quality in a CASIC Environment. *Proceedings of American Statistical Association: Survey Research Methods Section*, 41-49. Retrieved from <http://ww2.amstat.org/sections/srms/Proceedings/>
- Couper, M. P. (2017). Birth and Diffusion of the Concept of Paradata. *Advances in Social Research*, 18, 14-26.
Retrieved from <http://www.kyoto-info.com/kyoto/books/socialresearch/18.html> in Japanese (translated by W. Matsumoto), English version available at: http://jasr.or.jp/english/JASR_Birth%20and%20Diffusion%20of%20the%20Concept%20of%20Paradata.pdf
- Couper, M. P., & Singer, E. (2013). Informed Consent for Web Paradata Use. *Survey Research Methods*, 7(1), 57-67. doi:10.18148/srm/2013.v7i1.5138
- ESOMAR/GRBN (2015). *Guideline on Online Research*. Retrieved from <https://www.esomar.org/what-we-do/code-guidelines/esomargrbn-online-research-guideline>
- ESOMAR/GRBN (2017). *Guideline on Mobile Research*. Retrieved from <https://www.esomar.org/what-we-do/code-guidelines/esomargrbn-guideline-on-mobile-research>
- Felderer, B., & Blom, A. G. (2019). Acceptance of the Automated Online Collection of Geographical Information. *Sociological Methods & Research*, online first. doi:10.1177/0049124119882480
- Heerwegh, D. (2003). Explaining Response Latencies and Changing Answers Using Client-side Paradata from a Web Survey. *Social Science Computer Review*, 21(3), 360-373. doi:10.1177/0894439303253985
- Kaczmarek, L., & Neubarth, W. (2007). Nicht-reaktive Datenerhebung: Teilnahmeverhalten bei Befragungen mit Parادات evaluieren. In M. Welker & O. Wenzel (Eds.), *Online-Forschung 2007. Grundlagen und Fallstudien* (pp. 293-311). Köln: Herbert von Halem Verlag.
- Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to Participate in Passive Mobile Data Collection. *Public Opinion Quarterly*, 83(S1), 210-235. doi:10.1093/poq/nfz007
- Kreuter, F. (2013). Improving Surveys with Paradata: Introduction. In F. Kreuter (Ed.), *Improving Surveys with Paradata. Analytic Uses of Process Information* (pp. 1-9). Hoboken, NJ: John Wiley & Sons.
- Kunz, T., & Gummer, T. (2020). Understanding Respondents' Attitudes Toward Web Paradata Use. *Social Science Computer Review*, 38(6), 739-753. doi:10.1177/0894439319826904
- Kunz, T., & Hadler, P. (2020). Collection and Use of Web Paradata. *Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines)*. doi:10.15465/gesis-sg_en_037
- Kunz, T., Landesvatter, C., & Gummer, T. (2020). Informed Consent for Paradata Use in Web Surveys. *International Journal of Market Research*, 62(4), 396-408. doi:10.1177/1470785320931669
- RatSWD (2020). *Handreichung Datenschutz*. 2nd fully revised edition. RatSWD Output 8(6). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). Retrieved from <https://www.ratswd.de/publikation/output-series/0806>
- Revilla, M. A., Couper, M. P., & Ochoa, C. (2019). Willingness of Online Panelists to Perform Additional Tasks. *methods, data, analyses*, 13(2), 223-252. doi:10.12758/mda.2018.01
- Sakshaug, J. W., Schmucker, A., Kreuter, F., Couper, M. P., & Singer, E. (2016). Evaluating Active (Opt-in) and Passive (Opt-out) Consent Bias in the Transfer of Federal Contact Data to a

- Third-Party Survey Agency. *Journal of Survey Statistics and Methodology*, 4, 382-416.
doi:10.1093/jssam/smw020
- Sattelberger, S. (2015). *How the Timing of Informed Consent on Paradata Use Affects Response Behavior*. Paper presented at the GOR, Köln.
- Schaar, K. (2017). *Die informierte Einwilligung als Voraussetzung für die (Nach-)Nutzung von Forschungsdaten*. Retrieved from https://www.ratswd.de/dl/RatSWD_WP_264.pdf
- Schlosser, S., & Höhne, J. K. (2020). ECSP - Embedded Client Side Paradata. Retrieved from <https://zenodo.org/record/3782592#.XuxrAmgzY2w>